Australian Bureau of Statistics

2062.0 - Census Data Enhancement project: An update, 2011
Latest ISSUE Released at 11:30 AM (CANBERRA TIME) 28/11/2014 Ceased

Summary

Summary - Key outcomes

SUMMARY

Census Data Enhancement continues to be a key focus for the ABS. For information on how 2016 Census data is being used in data integration projects please refer to Information Paper: Census of Population and housing - Products and Services, 2016 (cat. no. 2011.0.55.001).

The Census Data Enhancement (CDE) project is a major ABS initiative that integrates data from the Census of Population and Housing with other datasets to efficiently create new, enhanced data sources for statistical, policy and research purposes. This project also adds value by bringing Census data together across time. The CDE project fits within ABS' broader Data Integration program.

KEY OUTCOMES OF THE CENSUS DATA ENHANCEMENT (CDE) PROJECT, 2011

- Release of the first Australian Census Longitudinal Dataset (ACLD) to provide a richer statistical view of Australian society, now and into the future. This dataset (formerly referred to as the SLCD), contains a 5% random sample of records from the 2006 Census linked with records in the 2011 Census. This allows for enhanced comparison over time between Censuses and the potential for linking longitudinal Census data with other data sets.

- Improved evidence-base for informed decision-making through initiatives such as the Indigenous Mortality Project, the creation of the Australian Census and Migrants Integrated Dataset (ACMID) and the 2011 Census to Vocational Education and Training in Schools dataset. These initiatives have improved the quality and availability of statistical evidence in areas of significant government and societal interest, such as COAG's Closing the Gap on Indigenous Disadvantage. The creation and dissemination of further statistical outputs will be a focus in the future.

- Promotion of the CDE integrated datasets through such initiatives as training, presentations at conferences, client initiated talks and guiding material.

- Implementation of new ABS governance mechanisms (in line with the Commonwealth's High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes) to ensure ABS' high standards of data secrecy, security, privacy and transparency are maintained.

This publication provides further information on the outcomes of the 2011 CDE project, as

well as links to the CDE studies, datasets and relevant information.

IN BRIEF - 2011 CDE DATA INTEGRATION OUTCOMES

Statistical initiatives	Status
Indigenous Mortality Study (to underpin Life expectancy estimates and Life tables)	released
Australian Census Longitudinal Dataset (ACLD)	released
Australian Census and Migrants Integrated dataset (ACMID)	released
Outcomes from Vocational Education and Training (VET) in Schools, experimental estimates (2011 Census, 2006 VET)	released
Educational outcomes, experimental estimates, Tasmania, (2011 Census, Government school enrolments and NAPLAN)	released
Mental Health Services - Census Data Integration Project	released
Australian Early Development Index - Census Data Integration Project	in progress
Australian Government's Settlements Database (SDB) and the ACLD	under consideration
Australian Government's Social Security and Related Information (SSRI) and the ACLD	under consideration

Quality studies	Status
Census to Census Dress Rehearsal (2011 to 2010)	completed
Migrants Quality Study (2011 Census)	released
Education Quality Study (2011 Census)	released

Outcomes of the 2011 Census Data Enhancement project

OUTCOMES OF THE 2011 CENSUS DATA ENHANCEMENT PROJECT

The overarching objective of the Census Data Enhancement (CDE) project is to enhance the relevance and potential of ABS Census data and other official datasets. Integrating Census data with other datasets is an efficient and effective way of creating new data from already existing sources to address important questions about Australian society.

As previewed in the 2010 release of *Census Data Enhancement Project: An Update,* the ABS has produced a range of data integration outputs using Census and other data. An overview of each of these outputs is presented in this section. Further detail for each project can be found in the source publications listed on the Related Information tab

STATISTICAL INITIATIVES

Statistical outputs from the CDE project provide new insights into areas and groups of interest within Australian society, such as education, migrants and Aboriginal and Torres Strait Islander peoples. These initiatives contribute to a richer statistical view of Australian society and an improved evidence base for decision making. They make new data available to the community, researchers and policy makers, in a cost effective way, without increasing the collection of information.

The 2011 CDE project has demonstrated that data integration can deliver:

- new data to address important areas of interest not able to be answered from existing sources; and
- improved understanding and accuracy of available information by comparing data from different sources.

The main statistical outputs foreshadowed in 2010 were the compilation of improved life expectancy estimates for the Aboriginal and Torres Strait Islander population, the first Australian Census Longitudinal Dataset (ACLD); and the linking of Census data (including longitudinal Census data) more generally with other data sources. These are described in the following sections.

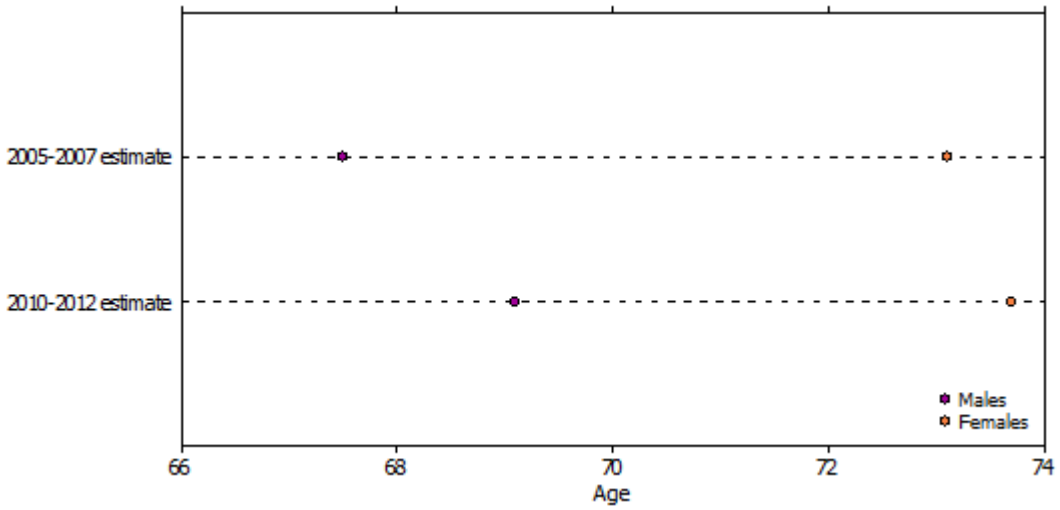
1. Indigenous Mortality Project - Death registrations and 2011 Census

The aim of the Indigenous Mortality Project is to improve the accuracy of life expectancy estimates and other mortality data for the Aboriginal and Torres Strait Islander population. This project, used extensively by the Council of Australian Governments, contributes to high quality data to inform reporting on the Australian Government's overall aim of closing the life expectancy gap between Aboriginal and Torres Strait Islander peoples and other Australians.

Using the highest quality linkage methodology available, the ABS linked 2011 Census records to deaths which occurred in the year after the Census. This was done to accurately assess the consistency of Indigenous status across the two datasets and built on work undertaken following the 2006 Census. The results were used to compile and improve the accuracy of Aboriginal and Torres Strait Islander life tables and life expectancy estimates. More broadly the project provides strategies for improving Indigenous identification in administrative data.

Graph 1 shows estimates of Indigenous life expectancy at birth for men and women for the years 2005-07 and 2010-12. Using the 2012 method of estimation, life expectancy for Aboriginal and Torres Strait Islander men increased from 67.5 years in 2005-07 to 69.1 years in 2010-12, and for women from 73.1 years in 2005-07 to 73.7 years in 2010-12.

Graph 1 - LIFE EXPECTANCY AT BIRTH FOR ABORIGINAL AND TORRES STRAIT ISLANDER AUSTRALIANS(a)



(a) 2005-07 estimates revised using the 2010-12 method to enable effective comparisons over time

Source: Life Tables for Aboriginal and Torres Strait Islander Australians

Further information and results can be found in Information Paper: Death Registrations to Census linkage project - Key findings for Aboriginal and Torres Strait Islander Peoples, 2011-2012 (ABS cat. no. 3302.0.55.005)

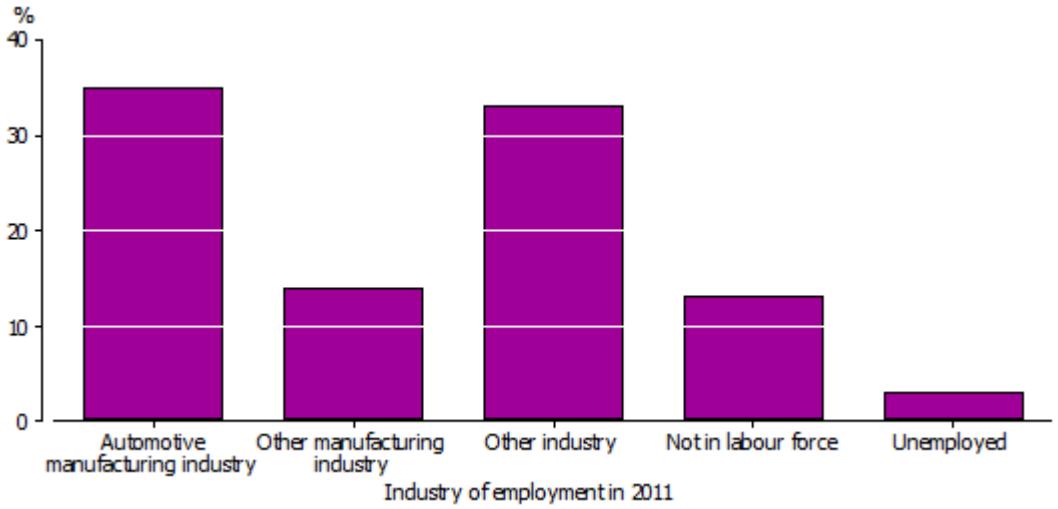
2. Creation of the Australian Census Longitudinal Dataset

First released in 2013, the Australian Census Longitudinal Dataset (ACLD) is Australia's largest longitudinal dataset bringing together a 5% random sample of around one million records from the 2006 Census with records from the 2011 Census. The ACLD will provide insight into Australians' journeys through life as successive Censuses are added. The ACLD provides a unique opportunity for researchers and policy makers to examine pathways and transitions of population groups.

The Australian Government Department of Industry used the Australian Census Longitudinal Dataset to look at the possible impact of car manufacturing closures in Victoria and South Australia. To do this they examined the transitions of workers out of the automotive industry between 2006 and 2011 (shown in Graph 2).

Graph 2 shows the industry of employment in 2011, for people who were in the automotive industry in 2006. It shows that of these people, 35% still worked in the Automotive manufacturing industry, while a third were employed in other industries and 16% were unemployed or not in the labour force.

Graph 2 - INDUSTRY OF EMPLOYMENT IN 2011 FOR PEOPLE WHO WERE IN THE AUTOMOTIVE MANUFACTURING INDUSTRY IN 2006



Source: The Australian Census Longitudinal Dataset

Further information can be found in Information Paper: Australian Census Longitudinal Dataset, Methodology and Quality Assessment (ABS cat. no. 2080.5).

The ACLD microdata product is available via the ABS website at Microdata: Australian Census Longitudinal Dataset (cat. no. 2080.0).

3. Bringing together the Australian Census Longitudinal Dataset with other datasets for statistical and research purposes

Combining the Australian Census Longitudinal Dataset (ACLD) with large administrative datasets has the potential to significantly enhance the statistical value of both datasets. To illustrate, a project linking the ACLD to the Australian Government's Settlement Database (SDB) is under consideration. This project would support longitudinal analysis of factors that impact migrants settling in Australia.

Consideration is also being given to linking the ACLD (and the 2011 Census as a whole) with Social Security and Related Information (SSRI) data to enable analysis of the characteristics and pathways of people in receipt of social security payments. The enhanced information would contribute to the evidence base for decision-making in this important social policy area.

4. Bringing together 2011 Census data with other datasets

One of the objectives of the 2011 CDE project is to use the linkage methodology developed in CDE quality studies (listed below) to inform a broader program of data integration between the Census and administrative datasets. This enables additional projects to be undertaken where there is both demand and statistical value.

4.1 Australian Census and Migrants Integrated Dataset

The Australian Census and Migrants Dataset (ACMID) aims to provide an evidence base to understand outcomes for migrants, such as the relationship between visa class (from the administrative data) and their post-arrival social and economic experiences (from Census data). The new enhanced dataset contributes to the evidence base of migrant data available in Australia and assists with development and evaluation of immigration policies and programs.

The ACMID combines administrative information from the Australian Government's Settlement Database (SDB) on visa class, application status and onshore/offshore processing with variables from the 2011 Census. The dataset was informed by the work done in the 2011 Migrants Quality Study (See Quality Studies below).

The ACMID provides a range of valuable insights on migrant outcomes that are not available in the administrative or Census datasets.

Graph 3 shows the personal weekly income of permanent migrants aged 15 and over by their visa class (skilled, family and humanitarian). Skilled migrants generally have a higher weekly incomes than those in the Family and Humanitarian streams. Nearly three quarters of Humanitarian stream migrants had incomes of less that $600 a week. Two fifths of Family stream migrants were in the lowest income group ($1 - $299) or had no income.

Graph 3 PERSONAL WEEKLY INCOME OF PERMANENT MIGRANTS BY VISA STREAM, 15 YEARS AND OVER - 2011

Source: Australian Census and Migrants Integrated Dataset

Further information and results can be found in Understanding Migrant Outcomes - Enhancing the Value of Census Data, 2011 (ABS cat. no. 3417.0).

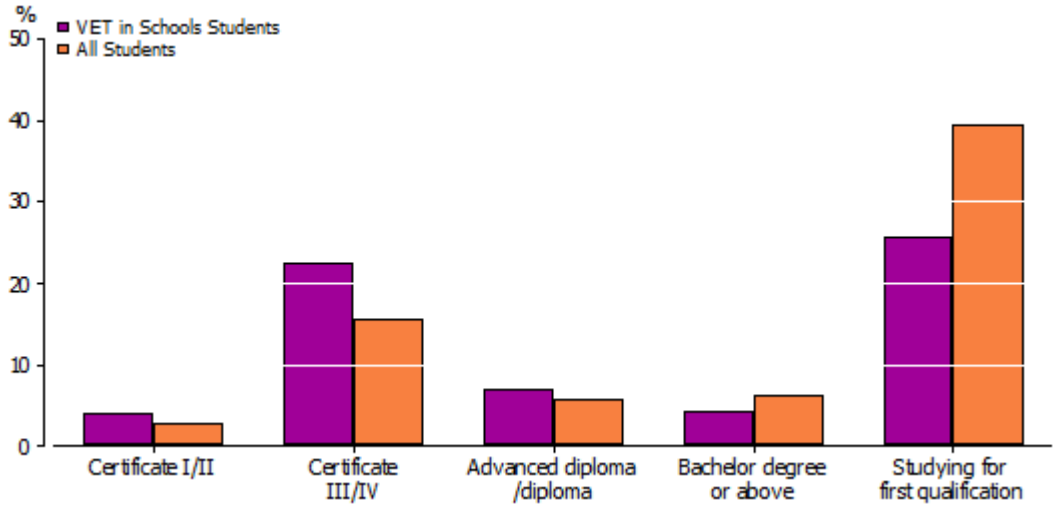
The ACMID 2011 TableBuilder product is available via the ABS website at Microdata: Australian Census and Migrants Integrated Dataset, 2011 (ABS cat. no. 3417.0.55.001)

4.2 2011 Census and Vocational Education and Training in Schools dataset

The combination of Vocational Education and Training (VET) in Schools and Census data allows examination of the characteristics and post-school outcomes of students who undertake a course of study under the VET in Schools program. It provides insights into year 12 completion rates, engagement and employment outcomes of Aboriginal and Torres Strait Islander students and the employment outcomes of students in trade related fields of study. Importantly the analysis can be undertaken for small populations, small geographies and by fine qualification type and by fine level of industry of employment.

Graph 4 shows the proportion of students in Year 11 in 2006 who have a qualification or who are studying for a qualification in 2011. More VET in schools students had completed a non-school qualification (41%) by 2011, compared with students overall (just under one third). Around one quarter of VET in Schools students were studying for their first qualification in 2011.

Graph 4 PROPORTION OF STUDENTS(a) WITH A QUALIFICATION(b) OR STUDYING IN 2011



a) Year 11 students in 2006
b) Qualifications completed by 2011
Source: Outcomes from Vocational Education and Training in Schools, experimental estimates

Further information and results can be found in Outcomes from Vocational Education and Training in Schools, experimental estimates, Australia, 2006-2011 (ABS cat. no. 4260.0)

4.3 2011 Census and Tasmanian Government school enrolment records and NAPLAN data

This project enabled analysis into the socioeconomic context of student achievement in Tasmania, and the destination and outcomes of Tasmanian early school leavers and Year 12 graduates. Tasmanian Government school enrolment records, along with National Assessment Program - Literacy and Numeracy (NAPLAN) data were linked to the 2011 Census to enhance the evidence base available to examine the socioeconomic context of student achievements, as well as the outcomes of young people after they leave school.

Further information and results can be found in Education outcomes, experimental estimates, Tasmania, 2006-2013 (ABS cat. no. 4261.6)

4.4 2011 Census and Australian Early Development Index dataset

This project links 2011 Census data, including family characteristics, with information on early childhood from the Australian Early Development Index to create a dataset to provide information for policy and program development in the area of early childhood education. This project is currently in progress.

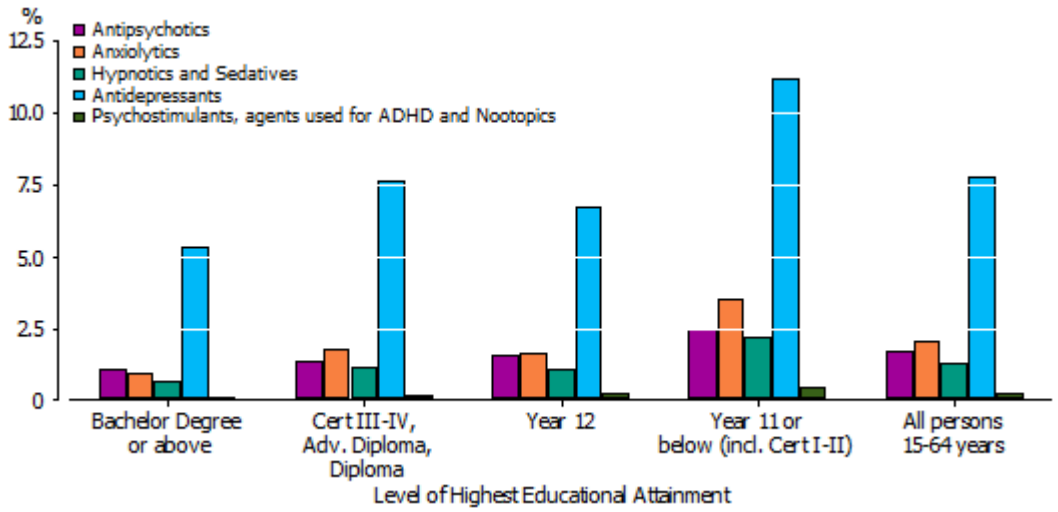
4.5 Mental Health Services - Census Data Integration Project

This project was undertaken to inform the National Review of Mental Health Services and Programs. The Mental Health Services-Census Integrated Dataset 2011, integrates the 2011 Census with subsidised mental health-related items from the Medicare Benefits Schedule (MBS) and the Pharmaceutical Benefits Scheme (PBS) from 2011. This integrated dataset will contribute significantly to the understanding of the characteristics of people who were accessing these subsidised services. An example of the information that is able to be extracted is provided in Graph 5.

Graph 5 shows the proportion of people (aged from 15 to 64) who accessed subsidised mental health-related prescription medication in 2011. This population is shown by their

highest level of educational attainment and type of drug accessed (antipsychotics, anxiolytics, hypnotics and sedatives, antidepressants, psychostimulants). Anitdepressants were the most commonly used medication across all levels of educational attainment. In 2011, 6.4% of Australians whose highest level of qualification was a Bachelor degree or higher, accessed a PBS subsidised mental health medication compared with 14.5% of those whose highest level of education was Year 11 or below.

Graph 5 PROPORTION OF AUSTRALIAN POPULATION (AGED 15-64) WHO ACCESSED PBS SUBSIDISED MENTAL HEALTH-RELATED PRESCRIPTION MEDICATION - 2011



Source: Characteristics of people using mental health services and prescription medication

Further information and results can be found in Characteristics of people using mental health services and prescription medication, 2011 (ABS cat. no. 4329.0)

5. Other projects

ABS is keen to pursue further opportunities for data linkage projects using the 2011 Census, including linking other datasets with the ACLD. For more information about linkage opportunities email the Director, Data Linkage Centre, Australian Bureau of Statistics at data.integration@abs.gov.au.

The ABS Privacy Policy outlines how the ABS handles any personal information that you provide us.

QUALITY STUDIES

Census Data Enhancement (CDE) quality studies are undertaken to:

- assess different linking strategies and test their quality; and
- improve data integration processes for future projects.

CDE quality studies have compared linkage strategies that use a range of available variables to link datasets with those that use name and address. The studies have concluded that linking without using name and address generally gives good quality and highly representative datasets for analysis. They have also shown that some groups within

the community are more difficult to link than others (for example, people who are highly mobile) pointing to the need for further research into linkage methodology.

The CDE quality studies form the foundation for the use of data integration to produce national statistics.

Four data integration quality studies were proposed in 2010. The outcomes of these are described below.

2011 Census to 2010 Census Dress Rehearsal

This quality study was undertaken to link data from the 2010 Census dress rehearsal to the 2011 Census to investigate the quality of different linkage strategies. The findings were used to revise and refine the linking strategy for the Australian Census Longitudinal Dataset (see 2. Creation of the Australian Census Longitudinal Dataset)

The 2011 Migrants Quality Study

The purpose of this quality study was to evaluate the dataset created by bringing together the Australian Government's Settlement Database (SDB) with the 2011 Census. The study focussed on assessing improvements to the linkage strategy, flowing from the corresponding 2006 feasibility study. The 2011 findings enabled an improved linkage methodology which was used to create the Australian Census and Migrants Integrated Dataset (see 4.1 Australian Census and Migrants Integrated Dataset). The report of this study can be found at Assessing the Quality of Linking Migrant Settlement Records to 2011 Census Data (ABS cat. no. 1351.0.55.043).

The Education Quality Study - linking school enrolment records to 2011 Census data

This quality study linked 2011 Census data with 2010 and 2011 government school enrolment records from Queensland, South Australia, Tasmania and the Northern Territory. The results indicated that linking government school enrolment records, and education records more generally, to the Census could be undertaken effectively and safely. It produced recommendations for ways of proceeding with education data linkage. The report of this study can be found at Research Paper: Assessing the Quality of Linking School Enrolment Records to 2011 Census Data (ABS cat. no. 1351.0.55.041).

2011 Census to a Western Australian Enhancement Mortality dataset

After discussions with stakeholders it was decided that this quality study would not be undertaken.

Preface

PREFACE

The Australian Census of Population and Housing is one of Australia's key statistical datasets underpinning a number of essential statistical assets presenting the most detailed view available of Australian society. The ABS Census Data Enhancement (CDE) project aims to realise the potential of this powerful dataset even further, bringing it together with existing data sources to unlock new insights into Australian society. This paper presents outcomes from the 2011 CDE project.

Building on exploratory work associated with the 2006 Census, the 2011 CDE project has successfully expanded the range of statistics available for research and policy purposes. It has done this by linking the Census to high value data sources to allow more in-depth analysis than would be possible using the datasets in isolation. The creation of the first Australian Census Longitudinal Dataset has been a highlight. It provides a unique opportunity for researchers and policy makers to examine pathways and transitions of population groups. Linking Census data with death registration, migration and education datasets has also led to improvements in the evidence base for decision making, as well as data quality and availability.

I would like to thank the Australian community, governments and organisations for their support. Together we have been able to improve and efficiently expand the range of official statistics available to Australian society whilst continuing to maintain the highest standards of confidentiality and privacy.

Jonathan Palmer
Acting Australian Statistician
November 2014

Introduction

INTRODUCTION

BACKGROUND

In August 2005 following extensive consultation and discussion, the Australian Statistician announced his intention to proceed with the Census Data Enhancement (CDE) project. The project was first undertaken for the 2006 Census, and has continued on for the 2011 Census. This paper presents outcomes from the 2011 CDE project.

OBJECTIVES

The overarching objective of the Census Data Enhancement (CDE) project is to enhance the relevance and potential of ABS Census data and other official datasets. The project aims to create new data sources for statistical, policy and research purposes, by bringing together existing data from the Census of Population and Housing with other data sources. A key aim of the project is to link Census data over time. This was achieved in 2013 with the release of the Australian Census Longitudinal Dataset (formerly referred to as the Statistical Longitudinal Census Dataset).

As well as these objectives, the CDE project also benefits the public by providing:

- an improved and expanded range of official statistics;
- a cost-effective and efficient way of leveraging existing data to meet current and emerging information requirements, without increasing respondent load; and
- the opportunity to support policy development, research and discussion more effectively in areas such as societal well-being, economic resources, education and health.

AUTHORITY

The CDE project is consistent with the legislated function of the ABS to maximise the use, for statistical purposes, of information available to official bodies.

The ABS Data Integration program, including the CDE project, accords with ABS legislated functions to:

- collect, compile, analyse and disseminate statistics and related information;
- avoid duplication of collection by official bodies, of information for statistical purposes; and
- achieve maximum possible utilisation, for statistical purposes, of information available to official bodies.

To respect the privacy of individuals and to protect the confidentiality of their data, the ABS operates under the *Census and Statistics Act 1905* and complies with the *Privacy Act 1988*. The use and release of data from the CDE project is governed by the provisions outlined under both these Acts. As an accredited Integrating Authority, the ABS also abides by the High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes.

Further information about Statistical Data Integration can be found via the ABS website.

Looking ahead

LOOKING AHEAD

- The ABS will continue to undertake linkage projects as part of the 2011 Census Data Enhancement (CDE) project where there is demand, public good, statistical value and available resources. As described in this paper, linking data with the Census (including the longitudinal Census dataset) has contributed to understanding social transitions and experiences, created enhanced data sources and helped to improve data accuracy.

- The statistical outputs of the 2011 CDE project have demonstrated the potential of data integration to replace direct collection, to develop new datasets (including longitudinal datasets) and to improve the quality of key estimates derived from administrative sources. The creation of the Australian Census and Migrants Integrated Dataset (ACMID), the 2011 Census to Vocational Education and Training in Schools dataset, the Australian Census Longitudinal Dataset (ACLD) and the work of the Indigenous Mortality Project, have efficiently enhanced the relevance of existing and new datasets whilst safeguarding the privacy of individuals. The CDE outputs have also improved the evidence-base of Australian official statistics for policy and research.

- Data integration will continue to be a central element of successive Censuses and is an increasingly important element of the broader ABS work program. This will be particularly important as the ABS seeks to continue to fulfil its official statistical role in an efficient and effective manner.

Confidentiality and Privacy

CONFIDENTIALITY AND PRIVACY

OVERVIEW

The 2011 Census Data Enhancement (CDE) project has been undertaken with privacy and confidentiality at the centre of all stages of the project. There have been no breaches of personal identification or privacy.

All personal information used in the CDE project is kept confidential. The *Census and Statistics Act 1905* guarantees this protection and legally prevents all ABS staff (including temporary employees) from disclosing information in a manner that is likely to enable the identification of a person or organisation. In addition, the ABS undertakes its operations in accordance with the *Privacy Act 1988.*

This section describes the processes for managing personal information in the CDE project. These include:

- legislative protection;
- destruction of Census forms and name and address information;
- functional separation;
- access to ABS information;
- data security; and
- accreditation as a Commonwealth Integrating Authority.

LEGISLATIVE PROTECTION

All ABS officers are bound by strict secrecy provisions under the *Census and Statistics Act 1905*. Officers sign an undertaking of fidelity and secrecy to ensure that they are aware of their responsibilities about Census data. Section 19 of the *Census and Statistics Act 1905* forbids past or present ABS officers from divulging information collected under this Act, either directly or indirectly, under penalty of up to 120 penalty units (currently $20,400) or imprisonment for 2 years or both.

These protections apply to all data collected by, or supplied to the ABS, including the data used for the CDE project and to all datasets generated from these data.

More information about the legislation governing the ABS can be found at ABS Legislative Framework

DESTRUCTION OF CENSUS FORMS AND NAME AND ADDRESS INFORMATION

The ABS destroys Census forms (hardcopy and electronic) after statistical processing has been completed. Name and address information is not retained. The only exception is if a person explicitly requests that their data is to be archived, by answering the relevant question on the Census form to have their name-identified responses retained by the National Archives of Australia for release in 99 years time (see Glossary for further detail). The ABS does not retain copies of this information.

FUNCTIONAL SEPARATION

In accordance with the high level principles for data integration involving Commonwealth data, the ABS implements functional separation in all data integration projects. Functional separation means that staff undertaking data linkage projects are assigned to different roles so that no ABS officer can see the identifying details of an individual as well as the information they have provided, at the same time. Similarly, datasets containing personal information are stored separately and securely from datasets containing other information. These procedures ensure that an individual's identity remains protected during the linking and analysis process.

ACCESS TO CDE OUTPUTS

ABS confidentiality procedures ensure all aggregate and microdata outputs disseminated by the ABS are unlikely to enable the identification of a particular person.

Access to linked datasets is made available to researchers and policy makers through the TableBuilder subscription service. TableBuilder is an online tool that maintains confidentiality while allowing users to cross tabulate and analyse data. More information about this service can be found on the ABS website.

The ABS also offers consultancies to produce confidentialised, aggregate data tables, data cubes or publications specific to user needs. Standard data confidentiality rules apply. Contact data.integration@abs.gov.au for more information on linked datasets.

The ABS Privacy Policy outlines how the ABS handles any personal information that you provide us.

DATA SECURITY

The ABS maintains practices of a high standard to ensure the security of all information it holds. These include:

- strong security arrangements for all ABS information technology systems. ABS conforms with IT Security arrangements set out in the Australian Government Information Security Manual ASCI 33;
- strict control of access to all ABS premises in accordance with the Commonwealth Protective Security Manual to ensure compliance with legislative responsibilities;
- appropriate personnel security arrangements. Upon appointment all ABS staff undergo security checks and are required to sign an undertaking of fidelity and secrecy;
- a secured Internet gateway which is reviewed annually by Defence Signals Directorate;
- regular Protective Security risk reviews to ensure that security arrangements continue to be effective; and
- an ongoing program of security audits and reviews of computer systems and the physical environment.

ACCREDITED INTEGRATING AUTHORITY

The ABS is an accredited Integrating Authority, acknowledging that it has the requisite expertise, skills and knowledge, infrastructure and secure environment to undertake high

risk data integration projects involving Commonwealth data for statistical and research purposes. Accredited Integrating Authorities remain subject to all relevant legislation, such as the *Privacy Act 1988.*

Contact us

CONTACT US

We welcome feedback on the Census Data Enhancement (CDE) project. We would also like to hear from people interested in undertaking data linking projects involving Census or in finding out about data integration more generally.
Feedback on the CDE project can be sent to data.integration@abs.gov.au or addressed to:

The Director
ABS Data Linkage Centre
Australian Bureau of Statistics
Locked Bag 10
Belconnen
ACT 2616

The ABS Privacy Policy outlines how the ABS handles any personal information that you provide us.

About this Release

This publication summarises the outcomes of the 2011 Census Data Enhancement (CDE) project. It shows how the ABS has improved the evidence base for decision making and research by combining information from the Census of Population and Housing with other existing data sources.

History of Changes

16/06/2017

Updated Summary page to refer readers to new product, which will replace this publication, and Ceased flag added.

Explanatory Notes

Explanatory Notes

EXPLANATORY NOTES

ABBREVIATIONS

ABS Australian Bureau of Statistics
ACLD Australian Census Longitudinal Dataset
ACMID Australian Census and Migrants Integrated Dataset
AEDI Australian Early Development Index
CDE Census Data Enhancement
COAG Council of Australian Governments
NAPLAN National Assessment Program - Literacy and Numeracy
SDB Settlement Database
SLCD Statistical Longitudinal Census Dataset
SSRI Social Security and Related Information dataset
VET Vocational Education and Training

GLOSSARY

Australian Census Longitudinal Dataset

The Australian Census Longitudinal Dataset was created by linking a 5% random sample of individuals from the 2006 Census to records from the 2011 Census. During its development and prior to release, it was referred to as the Statistical Longitudinal Census Dataset.

Administrative dataset

Information (including personal information) collected by agencies for the administration of programs, policies or services (e.g. Medicare data, taxation data). Administrative data is one type of unit record level data.

Australian Early Development Index

The Australian Early Development Index measures how young Australian children are developing. The data is collected every three years and provides a population measure of children's development at the time they start primary school. The main aim of the AEDI is to provide data to help communities in the development and reorientation of services and systems to improve the health and wellbeing of young children.

Australian Government's Settlement Database

The Australian Government's Settlement Database contains statistical data from the administration of immigration programs. This includes overseas arrivals and departures data, where the period of duration is over 12 months, and visa data, including type of visa.

Census Dress Rehearsal

The Census Dress Rehearsal is generally conducted in the year prior to the Census of Population and Housing. It is used as a run-through of operational processes and data collection methods for the main Census, and is tested on a sample of dwellings across Australia.

Census processing period

The period of time immediately after the conduct of the Census of Population and Housing during which the Census forms are processed to produce statistical outputs. The Census

processing period has generally lasted 12 months.

Confidentialisation procedures

Confidentialising data involves altering a dataset to ensure that individual records can not be identified. Some examples of confidentialisation are:

- perturbation - changing the data slightly to reduce/remove the risk of disclosure, without significantly affecting aggregate results;
- combining and collapsing categories - combining several response categories into one, or reducing the amount of classificatory detail available in microdata; and
- suppression - not releasing information for unsafe cells, or deleting individual records or data items from the file.

Data

Data are measurements or observations that are collected as a source of information.

Data integration

Data integration involves bringing together multiple data sources, generally at the unit record level (i.e. for a person or organisation) or micro level (e.g. information for a small geographic area), to provide new datasets for statistical or research purposes. Data integration refers to the full range of management and governance practices around the process, including project approval, data transfer, linking and merging the data, and dissemination.

Data item

Any characteristic, number, or quantity that can be measured or counted.

Data linking

Data linking (also referred to as data linkage or record linkage) is one aspect of the data integration process. Data linking creates links between data from two or more sources based on common features present in those sources.

Dataset

A file containing the individual responses from a statistical collection, administrative records or register of information (for example disease register). Datasets are used to generate statistical output. A data set that has been formed through data linking is called a linked (or integrated) dataset.

De-identified data

De-identifying data involves two key steps:

1. De-identification of the data, which is removal of any direct identifiers (eg. name, address, Australian Business Number) from the data; and
2. Removing or altering any other information that may allow an individual to be identified, such as a rare characteristic of an individual, or a combination of unique or remarkable characteristics that enable identification.

Death register data

The registration of deaths is the responsibility of the individual State and Territory Registrars of Births, Deaths and Marriages and is based on the data provided on an information form. This information form is the basis of the data provided to the ABS for processing and production of death statistics.

Dissemination

Dissemination is the process of outputting data. Dissemination at the ABS can be achieved through publications, unit record files, and static tables, as well as through dissemination products such as TableBuilder and DataAnalyser.

Linkage methodology

The method used to link datasets.

Linkage quality

The quality of a linked dataset is evaluated through various measures. These include estimating the number of records that were linked correctly, examining the properties of the records not able to be linked, and assessing the under or over representation of population groups.

Linkage strategy

A linkage strategy is the approach taken for a particular data linking project. It takes into account the method of linking to be used, the level of accuracy required, the suitability and availability of linking variables, and the resources available.

Longitudinal dataset

A dataset which contains information for the same unit over a number of different points in time.

Medicare Benefits Schedule

Medicare Australia collects Medicare Benefits Schedule claims data under the *Medicare Act 1973.* This information includes type of medical service, cost and provider type for services received from GPs.

National Assessment Program - Literacy and Numeracy

NAPLAN is an annual national test held in literacy and numeracy for students in Years 3, 5, 7 and 9. All students across Australia undertake the same year level tests in the four domains: reading, writing, language conventions (spelling, grammar and punctuation), and numeracy.

Official statistics

Official statistics are defined as those statistics produced by government departments and agencies including statistics collected by surveys or from administrative systems.

Pharmaceutical Benefits Scheme data

The Department of Health collects information on purchases of medicines subsidised by the Australian Government through the Pharmaceutical Benefits Scheme. This information

includes type of medication, cost and prescriber type.

Privacy

In the context of data integration, privacy refers to the protection of an individual's personal information as defined by the Privacy Act. The *Privacy Act 1988* is an Australian law which regulates how personal information is collected, used, stored and disclosed.

Quality study

A quality study investigates the outcomes of linkages using a common set of input data sources but different linkage methods. Quality studies help to determine the feasibility of different linkage methods and identify areas for improvement.

Random sample

A method of sampling in which every unit in the population has a predetermined probability of being selected.

Statistical purposes

Functions related to the compilation, analysis and dissemination of statistics. Statistical purposes precludes use of a dataset for administrative or client management purposes, where there is an impact on specified individuals.

Statistical output

The result of any collection, storage, analysis and transformation of data where the individual statistical unit is of no interest in itself, and the results are presented in a form that does not reveal information about identifiable individuals.

Student enrolments data

Student enrolments data is collected by education departments in state and territory governments.

TableBuilder

TableBuilder is an ABS dissemination tool that confidentialises data on the fly, allowing users to undertake cross-tabulations.

Unit record data

Unit record data refers to data where each record represents observations for an individual or organisation. Unit record data may contain individual responses to questions on a survey questionnaire or administrative form. For example, a unit record would have one person's answers given to the questions 'In what year were you born?', 'what is your address?' and 'what is your employment status?'.

Variable

Any characteristic, number, or quantity that can be measured or counted.

Vocational Education and Training in Schools data

Data on VET in Schools are collected through administrative sources in each state and territory. These authorities submit the data to the National Centre for Vocational Education Research where a national dataset is compiled. Data are inclusive of all persons aged 15-19 years who are enrolled in a VET in Schools module or unit.